Stock Prediction using Machine Learning

Dr. J. Dhillipan¹, D. B. Shanmugam², Imran Quraishi³

¹Professor & Head, Department of Computer Applications, SRM IST, Ramapuram Campus, Chennai
²Assistant Professor, Department of Computer Applications, SRM IST, Ramapuram Campus, Chennai
³MCA Final Year, Department of Computer Applications, SRM IST, Ramapuram Campus, Chennai

Corresponding Author mail id: hod.mca.rmp@srmist.edu.in

Abstract — Stock trading is one of the foremost activities in the finance world. Stock market prediction is used to find the long run values of the stock and other financial factors influenced on a financial exchange. The technical and fundamental or the statistical analysis is employed by most of the stockbrokers while making the stock predictions. Python programming language in machine learning is used for the stock market prediction. In this paper, we have proposed a Machine Learning (ML) approach which trains from the available stocks data and gains intelligence and then uses the acquired knowledge for an accurate prediction. In stock market prediction, the aim is to predict the longer term value of the financial stocks of a corporation [1]. The recent trend in market prediction technologies is that the use of machine learning approach which makes predictions supported the values of current stock market indices by training on their previous values. Machine learning itself employs different models to form prediction easier and authentic. This paper focuses on Regression and Long Short Term Memory (LSTM) based Machine learning to predict stock values. The factors that are being considered include re-open, close, low, high and volume [2,3].

Keywords — Long Short Term Memory; LSTM; Tensorflow; Neural Network Module.

1. Introduction

The introduction of an intelligent computer application into stock prices prediction will effectively reduce skepticism and conservative options among many investments’ professionals and market participants. It will guide the investors in overcoming the uncertainty and imprecision in predicting the stock prices at the most times [4]. The characteristic in each stock markets have common uncertainty, imprecision which is said with their short and long-term future state. Prediction market has been a hot research area for several years. If any system which may consistently predict the trends of the dynamic stock exchange be developed, it might make the owner and therefore the users of the system wealthy. These data may be suffering from inflation or fluctuation of exchange rates especially in countries like Nigeria [5,6]. Back propagation neural network is usually used for price prediction [7]. This paper demonstrates the event of a neural network application for analyzing and predicting of stock exchange prices considering the technical and fundamental factors as our input against one factor employed by many papers reviewed. It is expected that the accuracy of the appliance are going to be high compared to similar applications.

2. Existing Methods

The existing system proposes that a company’s performance, in terms of its stock worth movement, is foreseen by internal communication patterns to urge early warning signals. We tend to believe that it is vital for patterns in company communication networks to be detected earlier for the prediction of great stock worth movement to avoid attainable adversities that a corporation could face within the stock exchange so as that stakeholders’ interests is protected the utmost amount as attainable. Support Vector Machine may be a machine learning technique utilized in recent studies to forecast stock prices [8,9]. This model tries to predict whether the stock price is going to increase or decrease within a period in future.

2.1 Drawbacks of Existing Methods

- Accuracy would decrease when setting more levels of stock exchange movement.
- The typical of prediction accuracies using Decision Tree because the classifier are 43.44%, 31.92%, and 12.06% for “two levels,” “three levels,” and “five levels,” respectively.
- These results indicate that the stock price is unpredictable while the traditional classifier is implemented.

3. Methodology

In this paper, we’ve a dataset containing stock prices of Google from January 2012 to December 2016 [10,11,12]. We are advised to use these stock prices for training the neural network and predict the stock prices for the month of January-2017. This is a Regression problem [13,14,15]. To achieve this goal, we trained a Recurrent Neural Network LSTM. We will use one of the deep learning libraries, Keras, to build the neural network.

DOI: 10.30726/ijlca/v8.i4.2021.84008
3.1 Feature Scaling

The performance of the neural network are going to be better if the whole training input is within the same range [16,17]. The stock prices aren't within the same range. So, we need to scale the training data such that they are in the same range. This process is called feature scaling. The two general methods for feature scaling are:

- **Standardization**
  \[ x' = \frac{x - \bar{x}}{\sigma} \]
  where, \( x \) is the original feature vector, \( \bar{x} \) is that the mean of that feature vector, and \( \sigma \) is its variance [18].

- **Normalization (Min-Max normalization)**
  \[ x' = \frac{x - \text{min}(x)}{\text{max}(x) - \text{min}(x)} \]
  where, \( x \) is an ingenious value, \( x' \) is that the normalized value.

It is always advised to use Normalization in the case of Recurrent Neural Network (RNN) [19]. Therefore, we use Min-Max normalization here.

The fit method only calculates the min and max values [20,21]. It doesn’t apply the formula on the training set. The fit transform method applies the min-max formula on the training set [22]. Once applying the min-max formula, the transformed values will be in the range 0 and 1 i.e., the training data (features) will be in the range 0 and 1.

3.2 Reshaping

The input shape of the LSTM is 3Dimensional tensor with shape (batch_size, timesteps, input_dim). Batch_size represents the number of iterations required to traverse through the entire training data. Timesteps represent the number of inputs required for each prediction. In our scenario it is 60.

If no activation function is specified in the dense layer then the linear activation is performed by default wherein, the dense layer takes a weighted sum of its inputs which corresponds to the predicted output.

3.3 Data Collection

A data set (or dataset) may be a collection of knowledge. In tabular data, a knowledge set corresponds to at least one or more database, where every column represents a specific variable and every row corresponds to a record of the info set in question. The dataset lists values for all variables, like height and weight of an object, for every member of the info set. Each value is known as a datum. Data sets also can contains a set of documents or files.
performance and quality of predictions, feature selection and spam tweets reduction are performed on the info sets.

- Moreover, we perform experiments to seek out such stock markets that are difficult to predict and people that are more influenced by social media and financial news. We compared results of various algorithms to seek out a uniform classifier. Finally, for achieving maximum prediction accuracy, deep learning is employed.

4. Results

We have seen that our model closely predicts the trend of the actual stock prices. The model can be further improved and experimented by considering the following (but not limited to):

- Training the model with more data. Eg: Here we have used 5 years of stock prices but you can train the model with 10 years of data.
- Increasing the number of time steps.
- Adding more LSTM layers.
- Increasing the units in the LSTM layer.
- Adding some other indicators. Eg: If you have the financial instinct that the stock price of some other companies might be correlated to the one of Google, you could add this other stock price as a replacement indicator within the training data.

5. Conclusion

Over the years, various machine learning techniques are utilized in stock exchange prediction, but with the increased amount of knowledge and expectation of more accurate prediction, the deep learning models which are being used nowadays have proven their advantage over traditional machine learning methods in terms of accuracy and speed of prediction. In this article, we'll discuss the Long-Short-Term Memory (LSTM) Recurrent Neural Network, one among the favored deep learning models, utilized in stock exchange prediction. In this task, we'll fetch the historical data of stock automatically using python libraries and fit the LSTM model on this data to predict the longer term prices of the stock. The prediction are often more accurate if the model will train with a greater number of knowledge set.
Moreover, within the case of prediction of varied shares, there could also be some scope of specific business analysis. We can study various pattern of the share price of various sectors and may analyse a graph with more different time span to fine tune the accuracy. This framework broadly helps in marketing research and prediction of growth of various companies in several time spans. Incorporating other parameters (e.g. investor sentiment, election outcome, geopolitical stability) that aren't directly correlated with the price may improve the prediction accuracy.

References


